Dear Client,

I’m writing to you in response to Sprocket Central Pty Ltd Datasets that you shared with our team.

* Transactions
* New Customer List
* Customer Demographic
* Customer Address

We have analysed the quality of the raw data and found several quality issues in the datasets.

|  |  |  |
| --- | --- | --- |
| **Datasets** | **Total Number of Records** | **Unique ID Records** |
| **Transactions** | 19446 | 19274 |
| **New Customer List** | 931 | 736 |
| **Customer Demographic** | 3415 | 2778 |
| **Customer Address** | 4000 | 4000 |

We believe that these issues needed to be addressed before we continue this further processing of the data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Datasets** | **Accuracy** | **Completeness** | **Relevancy** | **Validity** | **Consistency** |
| **Transactions** | **Profile:** missing | **Blanks** -Product ID, Online order | **Filtered Out**:  Cancelled -Order status | Price mentioned used to calculate Profits |  |
| **New Customer List** |  | **Blanks** - Job Title, Job Industry Category | **Format**: DOB | Calculated Age using DOB |  |
| **Customer Demographic** | **Age:** missing  **DOB:** inaccurate | **Blanks** - Job Title | **Format**: DOB  **Irrelevant**: Default column |  | **Inconsistent**: Gender |
| **Customer Address** |  |  |  |  | **Status:** Inconsistent |

Suggestions to mitigate data quality issues in future: -

1. **Accuracy Issues:**
2. **Customer Demographic Dataset**

Some calls in DOB column had inaccurate values and some of the cells in the Age column had missing and inaccurate values.

**Recommendation:** Filter out outliers in the DOB column and double check the Age and the Profit column to ensure there are no missing values.

1. **Completeness Issues:**
2. **Customer Demographic Dataset**

Some cells in the job\_title were blank.

1. **Transaction Dataset**

Both the online\_order and the brands were blank.

**Recommendation:** Double checking the columns to ensure that the columns don’t have blank spaces.

1. **Consistency Issues:**
2. **Customer Demographic Dataset**

Both the Female and Male genders were inconsistent.

1. **Customer Address Dataset**

The states were inconsistent.

**Recommendation:** Use selective options to minimize manual entry errors which in return will reduce human errors.

1. **Relevancy Issues:**
2. **Customer Demographic Dataset**

Irrelevant Default column.

1. **Transaction Dataset**

Irrelevant data values in the order\_status column.

**Recommendation:** Recheck and delete all the irrelevant data values.

1. **Validity:**
2. **Transaction Dataset**

Wrong data format in both the list\_price and product\_first\_solid\_date.

**Recommendation:** Ensure that column data values are in the same correct format.

Please look into the above mentioned quality issues along with the recommended changes. If all the suggestions are matched, then we can proceed with further analysis of data.